

# Augmenting XBRL Using UML: Improving Financial Analysis

Joseph H. Callaghan (E-mail: [callagha@oakland.edu](mailto:callagha@oakland.edu)), Oakland University

Arline Savage (E-mail: [savage@oakland.edu](mailto:savage@oakland.edu)), Oakland University

Vijayan Sugumaran (Email: [sugumara@oakland.edu](mailto:sugumara@oakland.edu)), Oakland University

## Abstract

*eXtensible Business Reporting Language (XBRL), a financial accounting application of XML, provides a taxonomy for facilitating the analysis of financial statement information. This taxonomy promises to enhance the financial analyst's ability to process financial information both cross-sectionally and temporally. However, limitations exist in the taxonomy that inhibit the analyst, intelligent agent or application to process information in the most effective way. In particular, financial statement element analysis can only be properly conducted when the additional semantics provided by additional disclosures are incorporated into the process. UML provides a conceptual framework to capture more meaningful semantics, particularly among related, collaborating objects. The purpose of this paper is to develop a framework that will couple the strengths of both technologies in order to develop better systems for financial statement analysis.*

## 1. Introduction

There is a growing level of excitement in the financial community around the eXtensible Business Reporting Language (XBRL) standard because of the expectation that it will make it easier to collect, aggregate and publish financial results from publicly held companies (Bovee et al. 2002; Coffin 2001; Zarowin and Harding 2000). According to XBRL.org, "The goal of XBRL is to provide an XML based framework that the global business information supply chain will use to create, exchange, and analyze financial reporting information including, but not limited to, regulatory filings such as annual and quarterly financial statements, general ledger information, and audit schedules." Thus, using XBRL, investment analysts can easily analyze and compare financial statements from different companies from an investment standpoint (Strand et al. 2001; Watson et al. 2000).

While this new standard facilitates the interoperability of financial data contained in various financial statements, the XBRL taxonomy tags footnotes and other disclosures in a way that does not reflect the natural way in which an analyst, intelligent agent or application works. For instance, before analyzing financial statement elements, an analyst would examine the auditor's opinion and refer to the general accounting policies disclosures, including e.g., inventory valuation and depreciation methods. In this way the quality and meaning of financial statement elements is enhanced. Similarly, when examining a particular financial statement element with a cross-reference to a footnote, the disclosure would add meaning by putting the element in context, elaborating it or disaggregating it. The lack of interaction between element tags and disclosure tags presents processing problems for applications using the XBRL document.

UML provides diagrams that permit the modeling of these natural interactions. These diagrams in turn can provide a schema for augmenting XBRL documents that then can be better processed by downstream users, including intelligent agents, and applications.

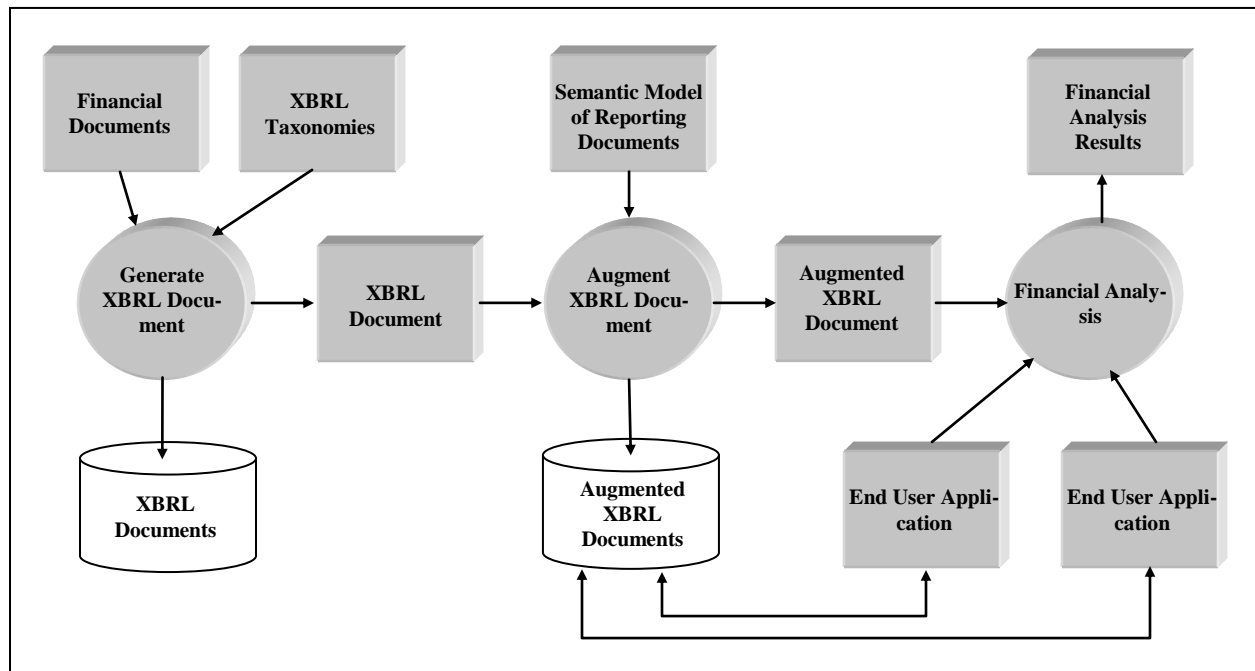
---

*Readers with comments or questions are encouraged to contact the authors via email.*

## 2. Framework

Figure 1 depicts the overall framework for the processing of financial information.

**Figure 1.**  
**Framework for Financial Processing Using Augmented XBRL Documents**



Financial documents, such as annual reports and SEC-mandated 10-K and 10-Q filings, provide the raw material for financial analysis. The order in which financial reports are normally presented differs, but current trends can be discerned from “Accounting Trends & Techniques,” published annually by the American Institute of Certified Public Accountants (AICPA), which reports the results of a survey of financial reporting trends garnered from the annual reports of 600 corporations registered with the Securities and Exchange Commission (SEC). In the latest survey, which is the one used for purposes of this study, the annual reports were those of selected industrial, merchandising, technology, and service companies across 46 industry classifications for fiscal periods ending between February 25, 2000 and February 4, 2001 (AICPA 2001).

Rule 14a-3 of the Securities Exchange Act of 1934 states that information required to be included in annual reports to stockholders should include audited balance sheets for the two most recent fiscal years, and income and cash flow statements for each of the three most recent fiscal years. In addition, the following should be included: (1) selected quarterly financial data; (2) disagreements with accountants on accounting and financial disclosure; (3) summary of selected financial data for last five years; (4) description of business activities; (5) segment information; (6) listing of company directors and executive officers; (7) market price of the company’s common stock for each quarterly period within the last two fiscal years; (8) management’s discussion and analysis of financial condition and results of operations; and (9) quantitative and qualitative disclosures about market risk (AICPA 2001, 2).

The major components of the financial reporting system, the order in which they typically appear, and the percentage of the occurrence (if available) according to the AICPA 2001 survey, are shown below:

1. Report of the Independent Accountants (or Audit Report) either follows the financial statements and notes (2001 – 49%), or precedes the financial statements and notes (2001 – also 49%). For the titles of the audit reports, 598 out of 600 companies, or 99.7%, included the words “independent” and “auditor” (AICPA 2001, 579).
2. Balance Sheet. All 600 annual reports contained comparative balance sheets, as required by the SEC Rule 14a-3. In the 2001 survey, 33% of the companies presented a balance sheet followed by an income statement. The predominant balance sheet format was the report form (84%), followed by the account form (16%). Only one company out of 600 used the financial position form. According to the AICPA (2001, 129), the “report form shows a downward sequence of either total assets minus total liabilities equal to stockholders’ equity or total assets equal to total liabilities plus stockholders’ equity,” while the “account form shows total assets on the left-hand side equal to the sum of liabilities and stockholders’ equity on the right-hand side.”
3. Income Statement. All 600 annual reports contained income statements for the three most recent fiscal years, as required by the SEC Rule 14a-3. In the 2001 survey, 55% of the companies presented an income statement followed by a balance sheet, while 4% presented an income statement first, followed by a statement of cash flows. Key words used in income statements titles were: income (47%); operations (33%); earnings (18%); and other words (2%) (AICPA 2001, 289). The income statements formats were: (1) single-step form with income tax shown as separate last item (22%), and (2) multi step form (78%), with 40% showing costs and expenses deducted from sales to show operating income, and 38% showing costs deducted from sales to give gross margin (AICPA 2001, 289).
4. Statement of Stockholders’ Equity, which included a Statement of Comprehensive Income in 81% of the sample cases that had reported comprehensive income (i.e., 519 companies out of 600). A separate statement of comprehensive income was included in 13% of the cases, while a combined attachment of income and comprehensive income was used for the remainder of the cases (6%). Of the sample of 600, 81 companies or 13.5% had no comprehensive income to report (AICPA 2001, 401).
5. Statement of Cash Flows. All 600 annual reports contained cash flow statements for the three most recent fiscal years, as required by the SEC Rule 14a-3. This was the final statement in 49% of the sample cases, and it followed the Income Statement and Balance Sheet in 47% of the cases. In 4% of the cases, it was put between the income statement and the balance sheet (AICPA 2001, 505). Of the 600 companies, 96.5 % used “cash flows” as part of the title, while the remainder used the singular, namely, “cash flow.”
6. Notes to the Financial Statements, the first of which is often the Accounting Policies disclosure. Accounting Principles Board (APB) Opinion No. 22 states that the preferable format is to present a Summary of Significant Accounting Policies preceding the Notes to the Financial Statements, or as the initial note, which would be Note 1.

The AICPA survey (2001, 39) found that SEC regulations (AICPA, 2001, 39) state the need for adequate disclosure. Financial statements alone cannot adequately provide all the information required, without notes to financial statements. The AICPA survey (2001, 39) lists examples the following commonly-used types of notes to financial statements: changes in accounting principles; retroactive adjustments; long-term lease agreements; assets subject to lien; preferred stock data; pension and retirement plans; restrictions on the availability of retained earnings for cash dividend purposes; contingencies and commitments; depreciation and depletion policies; stock option or stock purchase plans; consolidation policies; computation of earnings per share; subsequent events; quarterly data, segment information; and financial instruments. According to the most recent survey (AICPA 2001, 39), the manner in which the financial statements refer to the notes is as follows: (1) general reference only – 75%; (2) a combination of general and direct reference – 25%; direct reference only – a mere one company out of 600.

The nature of information most frequently disclosed in summaries of accounting policies, in descending order, is as follows (AICPA 2001, 39): depreciation methods (97%); consolidation policy (97%); use of estimates (96%); revenue recognition (92.5%); property (92%); inventory pricing (90%); cash equivalents (86%); amortization of intangibles (80%); impairment (71%); financial instruments (70.5%); earnings per share calculation (68%); inter-period tax allocation (67%); translation of foreign currency (66%); stock-based compensation (53%); nature of operations (52%); advertising costs (33%); research and development costs (30%); fiscal years (30%); credit risk concentrations (29%); employee benefits (27%); environmental costs (23%); and capitalization of interest (15%).

Although there is some structure to the SEC filings, the purpose of XBRL is to enhance these documents by developing tag sets for financial statement elements (Debreceeny and Gray 2001). For example, the following snippet provides a self-describing context for a particular financial statement element, *viz.* inventory:

```
<Asset date=12-31-2002>
  <Current Asset>
    <Inventory>10000</Inventory>
  </Current Asset>
</Asset>
```

This structure enhances the downstream processing of financial statement elements. If a set of companies used the same taxonomy at the same point in time, then cross-sectional analysis is facilitated. Similarly, if the same taxonomy is used over time for the same company, time-series analysis is facilitated. Finally, if both techniques are used panel data statistical analysis can be accomplished.

With respect to additional disclosures, the same inventory financial statement element would be referred to under General Accounting policies. For example:

```
<Note 1>      The company uses the FIFO method on a lower-of-cost-or-market basis for valuation of
               inventories...</Note 1>
```

Another disclosure, specifically cross-referenced to the inventory statement element might contain:

```
<Note 5>      The merchandise inventory amount contains $2,000 related to merchandise that had originally cost
               $3,000...</Note 5>
```

Thus, the meaning of the statement element is better described by the disclosure, both in terms of context (valuation method) and in terms of disaggregation (the \$1,000 markdown to lower-of-cost-or-market). Unfortunately, the current XBRL taxonomy does not permit the additional meaning to be systematically garnered. Thus, the value of the XBRL conversion is less than what could be attained with disclosure integration with statement element information.

UML diagrams could better represent the inherent inter-relatedness of these underlying objects. The modeling of these objects and their collaborations would better represent their semantics. A semantic model for reporting documents would enable the augmentation of XBRL documents such that downstream financial processing would be a better value-added proposition. In sum, modeling the more subtle relationships that exist within the financial statement documents and using those models provide a methodology for improving the downstream processing of the document content, whether by analysts, intelligent agents or applications.

### **3. UML-Based Modeling of Financial Documents**

Unified Modeling Language (UML) is the most commonly used modeling language today. Software engineers routinely use UML in systems development, particularly to specify the structure and behavior of the overall system. The primary benefits of UML are: a) easy transition from design to implementation, and b) UML specification can be executed to check for consistency and accuracy. UML unifies and formalizes many object-oriented approaches advocated by Booch et al. (1999) and Martin and Odell (1998). It provides a rich set of modeling constructs that can be used to capture the intricacies of the requirements of systems as well as the semantics of the application domain.

UML can be used to not only represent the structure and behavior of systems, but also to model various aspects of domain artifacts at a high level (Carlson 2001). For example, in the financial reporting domain, there are several types of reporting documents that exist with well-defined components and structures. These components are also interrelated, with specific processing implications. It is essential that these relationships are understood and captured to facilitate the automated processing of financial data contained in these documents. UML naturally lends itself in modeling the structural components of financial reporting documents, the relationships between them, and how these components are utilized in financial analysis and how they impact one another in various contexts. Thus,

modeling XBRL documents using UML would help enhance the semantic content of the document. This semantic augmentation can be exploited by the applications that use these documents for better financial analysis.

Overall, UML supports the following types of models (Booch et al. 1999): a) use case model – the specification of actions that a system or class can perform by interacting with outside actors, b) static model – describe the structure of components, semantics of data and messages in a conceptual way, c) dynamic model – captures the behavior of components, d) implementation model – describe the component distribution on different platforms, and e) object constraint language – formal language to express additional semantics within UML specification. Each model type supports one more diagrams. Some of these diagrams are listed in Table 1.

**Table 1.**  
**Common UML Model Types and Associated Diagrams**

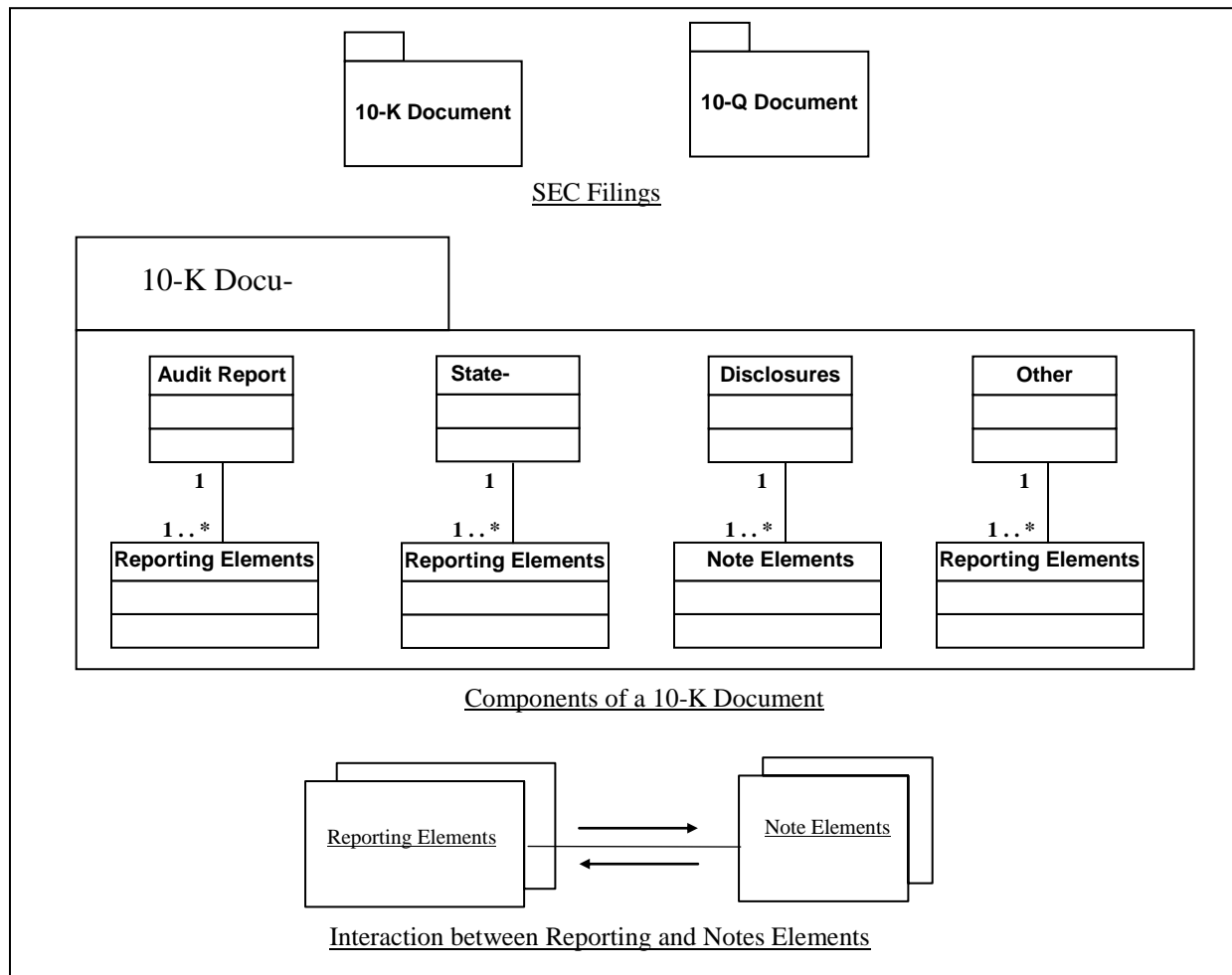
Model Type	Associated Diagrams
Use Case Model	Use case Diagrams
Static Model	Class Diagram Object Diagram Package Diagram
Dynamic Model	Collaboration Diagram Sequence Diagram Activity Diagram Statechart Diagram
Implementation Model	Component Diagram Deployment Diagram
Object Constraint Language	Constraints, Invariants Pre and Post Conditions Navigation Paths

We use a subset of the diagrams shown above in modeling an XBRL financial document. In particular, we use the class diagram and the package diagram to capture the structure of the financial document. The collaboration diagram is used to capture how the individual components of the document interact with each other, thus facilitating the integration of these components. Some of the elements of the financial document are used as input to certain types of financial processing activities, which could be expressed using the activity diagram. Figure 2 shows a high level overview of UML components for SEC filing documents. The interaction between the reporting elements and the disclosure elements is the key to understanding and using these filing documents.

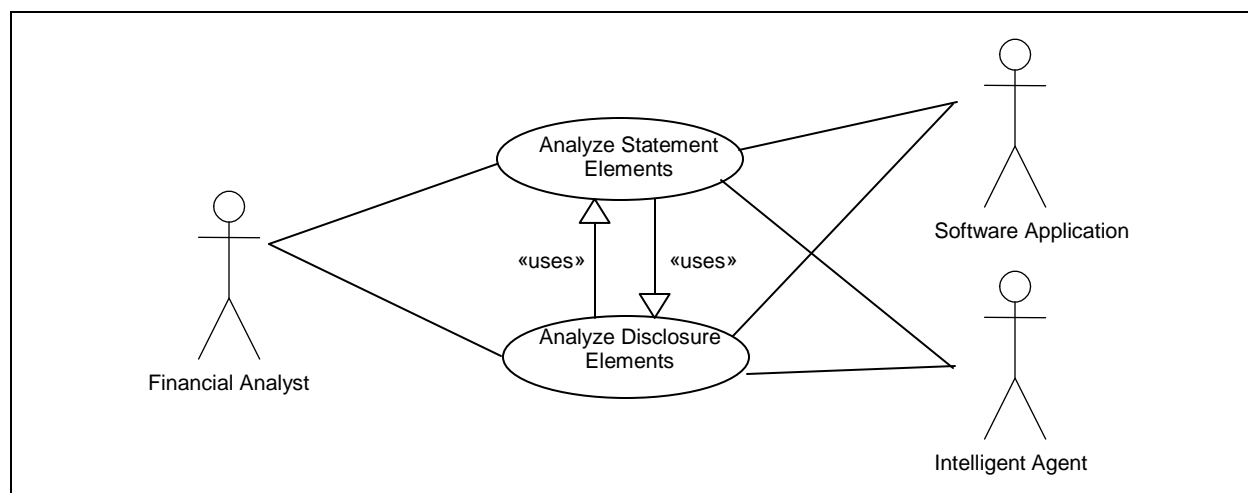
#### **4. Processing XBRL Documents**

Once the SEC filings are converted into XBRL documents, they can be utilized by various external entities that perform different types of financial analyses. The semantics of the XBRL documents improved through UML constructs, results in the effective use of the contents of these documents. In addition to human analysts, intelligent agents can also be employed in the analysis process. In particular, agent-based applications can be developed that can use the enhanced XBRL documents and automate the analysis process for a variety of financial computations. For example, standard ratio analysis may be performed, but on items from companies meeting certain disclosure requirements, e.g. only companies with clean opinions and LIFO inventories. Figure 3 shows a use case diagram that would serve as a starting point for developing software applications that can assist the analyst in processing financial statements.

**Figure 2.**  
**UML Components Representing SEC Filing Documents**




**Figure 3.**  
**Use Case Diagram for Analyzing SEC Filing Documents**



## 5. Conclusion and Suggestions for Future Research

This paper discusses the need for improving the semantics of SEC filing documents and presents a framework for augmenting the XBRL documents through the use of UML constructs. Such an enhanced document can be processed by human agents, as well as by software applications and intelligent agents. This results in the increased utilization of XBRL documents in financial analysis and decision-making.

Future work includes clearly articulating and modeling the various elements of disclosures and how they relate to financial statement elements, and developing a proof-of-concept prototype application. This framework for using UML to augment XBRL financial documents will provide richer semantics in the financial statement analysis domain. The UML schema could be used to generate code through forward engineering that would further process the XBRL document, making the augmented documents more amenable to further processing by financial analysts, intelligent agents and downstream software applications. Financial analysts could develop screening mechanisms and standard ratio analysis based upon the augmented documents. Intelligent agents could search for companies meeting conditions contained in the augmented documents. Software applications could further process the augmented documents performing standard analyses on a broader set of financial constructs. In general, the integration of the two technologies affords a synergistic method for enhanced financial statement analysis. 

## References

1. AICPA, "Accounting Trends & Techniques" (55<sup>th</sup> Ed.), AICPA, New York, NY, 2001.
2. Booch, G., J. Rumbaugh and I. Jacobson, *The Unified Modeling Language User Guide*, Addison Wesley, New York, NY, 1999.
3. Bovee, M., M. Ettredge, R.J. Srivastava and M. Vasarhelyi, "Enabling Semantic Consistency for Electronic Business Reporting: Does the Year 2000 Taxonomy Accommodate Current Business Financial Reporting Practice?" *American Accounting Association Information Systems Mid-Year Conference*, Orlando, FL, April 2002.
4. Carlson, D, *Modeling XML Applications with UML*, Addison Wesley, New York, NY, 2001.
5. Coffin, Z, "The Top Ten Effects of XBRL," *Strategic Finance*, pp. 64-67, June 2001.
6. Debreceeny, R. and G.L. Gray, "The Production and Use of Semantically Rich Accounting Reports on the Internet: XML and XBRL," *International Journal of Accounting Information Systems*, 2, pp. 47-74, 2001.
7. Martin, J. and J. Odell, *Object Oriented Methods: A Foundation* (UML Ed.), Prentice Hall, Englewood Cliffs, NJ, 1998.
8. Strand, C.A., B.L. McGuire, L.A. Watson and C. Hoffman, "The XBRL Potential," *Strategic Finance*, pp. 58-63, June 2001.
9. Watson, L.A., B.L. McGuire and E.E. Cohen, "Looking At Business Reports Through XBRL-Tinted Glasses," *Strategic Finance*, pp. 40-45, September 2000.
10. Zarowin, S. and W.E. Harding, "Finally, Business Talks the Same Language," *Journal of Accountancy*, pp. 24-30, August 2000.

---

**Notes**